**PSY 653 Module 12: Introduction to Multilevel Modeling**

**Gemma Wallace & Neil Yetz**

**April 29, 2020**

*Description of the dataset for the demo activity:*

*Description of the dataset for the try-it-yourself activity:*

\*(Datafile and description courtesy of Dr. Kim Henry)

A research team at a large University sought to determine if an 8 week summer program designed to encourage female high school students to pursue Data Science education and career paths was more effective if the program was team-focused vs. individual-focused. 500 females who were recruited to participate in the program completed an application packet that in- cluded a high school transcript, an online assessment of their current skills in math and computer science, and psychological as- sessments of self-efficacy for STEM disciplines. Using these data, the researchers created an index that binned the females into quintiles based on the likelihood of success in the summer program, the index ranged from 0 to 4, where 0 designated the highest likelihood of success and 4 designated the lowest likelihood of success. Once this “risk index” was created, one female from each quintile was randomly assigned to a team of 5 students such that each team had one female who had a risk index of 0, one who had a risk index of 1, and so forth. This created a total of 100 teams, each with 5 team members. Next, each team was randomly assigned to participate in either a team-focused version of the summer program, or an individual-focused version of the summer program. Following random assignment, the teams participated in the 8 week program. Throughout the program, a series of measures and assessments were collected. The data are in a file called mlm\_teams.csv, and below is a summary of the variables.

* team\_id: The team number, values of 1 to 50 denote team-focused teams (the treatment condition), and values of 51 to 60 denote individual-focused teams (the control condition).
* kid\_id: The personal ID number of the student.
* txcond: Condition indicator (0 = control condition, 1 = treatment condition).
* risk: The student’s risk quintile (ranges from 0 to 4, where 0 = lowest risk quintile, 4 = highest risk quintile)
* score: The student’s score on a final comprehensive exam to measure knowledge gained during the summer program. It ranges from 0 to 100, where a higher score denotes more knowledge.
* comafrd: At the end of week 4, all students completed a measure of belief in the communal affordances of a Data Science career (i.e., that a career in Data Science would afford the opportunity to reach communal goals, such as having a positive impact on society, developing close relationships with co-workers, altruism). The scale ranged from 1 to 9, where a higher score denoted a stronger belief that a career in Data Science would allow for communal goals to be met.

1. Create a new R notebook and load the following libraries: tidyverse, lme4, psych & MuMIn (Note: You will likely need to install the MuMIn package)
2. Read in the “mlm\_teams.csv” data.
3. Factor the team\_id variable.
4. Perform and interpret multilevel analysis using team\_id as a level-2 variable and risk as a level-1 variable to predict score.
5. Interpret the results, including the intraclass correlations.

**Try it Yourself Activity**

The data file “popularity.csv” is simulated data for 2000 pupils in 100 schools. Retrieved from: https://github.com/MultiLevelAnalysis/Datasets-third-edition-Multilevel-book/tree/master/chapter%202/popularity

* pupil - The personal ID number of the pupil.
* class - The class number
* Cextrav - pupil extraversion (10-point scale). Centered at the mean (0 = Mean value).
* popular - a popularity rating on a scale of 1–10 derived by a sociometric procedure.

**Demo Activity**

1. Read in the “popularity.csv” data.
2. Factor the class variable.
3. Perform and interpret multilevel analysis using class as a level-2 variable and Cextrav as a level-1 variable to predict popularity.
4. Interpret the results, including the intraclass correlations.